



EP0801344

Biblio

D sc

Claims

Page 1

Drawing

esp@cenet



An apparatus for reallocating logical to physical disk devices using a storage controller and method of the same

Patent Number: ☐ EP0801344, A3

Publication date: 1997-10-15

Inventor(s): YAMAMOTO YASUTOMO (JP); SATOH TAKAO (JP); YAMAMOTO AKIRA (JP)

Applicant(s):: HITACHI LTD (JP)

Requested Patent: ☐ JP9274544

Application Number: EP19970105448 19970402

Priority Number(s): JP19960085370 19960408

IPC Classification: G06F3/06

EC Classification: G06F3/06M, G06F11/10M, G06F11/20L

Equivalents: ☐ US5956750

Abstract

A storage controller (104) includes that it calculates an access frequency (500) of each logical disk (200); that it selects first logical disk device of which the access frequency exceeds a first predetermined value, the first logical disk device being allocated to a first physical disk device; that it selects a second logical disk device which has the access frequency equal to or less than a second predetermined value, the second logical disk device being allocated to a second physical disk device; and that it reallocates the first and second logical devices to the second

and the first physical disk devices, respectively.



Data supplied from the esp@cenet database - I2

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開 号

特開平9-274544

(43) 公開日 平成9年(1997)10月21日

(51) Int. Cl. ⁴	G 06 F 3/06	P 1	特許表示箇所
		G 06 F 3/06	540
			301 X
			302 E
			304 N
			320
	12/08	7623-5B	審査請求 未請求 請求項の数 9 O L (全 14 頁)

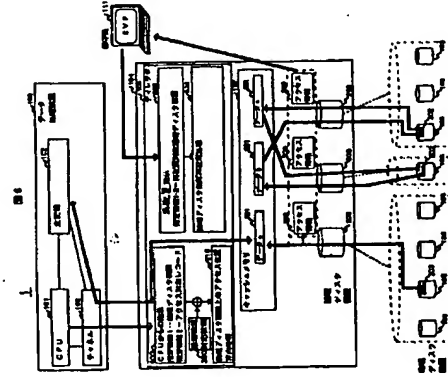
(21) 出願番号	特開平9-85370	(71) 出願人	00005108 株式会社日立製作所
(22) 出願日	平成8年(1996)4月8日	(72) 発明者	東京新千代田区神田豊河台四丁目6番地 山本 康文
		(72) 発明者	神奈川県横浜市港南区王禅寺1095番地 株 式会社日立製作所システム開発研究所内 山本 彰
		(72) 発明者	神奈川県川崎市麻生区王禅寺1095番地 株 式会社日立製作所システム開発研究所内 佐藤 孝夫
		(74) 代理人	神奈川県小田原市国府津2880番地 株式会 社日立製作所ストレージシステム事業部内 弁理士 有近 紳太郎

(54) 発明の名称 記憶制御装置

(57) 要約

【課題】 アクセス性能を向上する。
【解決手段】 各論理ディスク装置200に対するアクセス情報500を採取し、そのアクセス情報500をSV1111を通じて保守員に提示する。保守員の再配置情報620があると、指示された2つの論理ディスク装置200の間で物理ディスク装置105を位置替えし、全データを自動的に格納し直す。

【効果】 アクセス頻度の高い論理ディスク装置により高速な物理ディスク装置へ再配置することが出来る。シケンシャルアクセスの比率の高い論理ディスク装置よりシケンシャルアクセス性能の高い物理ディスク装置へ再配置することが出来る。



【特許請求の範囲】

【請求項1】 データ処理装置が直接アクセスを行う論理的記憶装置を複数にデータを記憶する物理的記憶装置に配置し、前記データ処理装置と前記物理的記憶装置とのデータ転送を制御する記憶制御装置において、予め定めた指標に基づいて前記論理的記憶装置を前記物理的記憶装置に再配置すると共に再配置先の物理的記憶装置にデータを連続的に格納する論理的記憶装置再配置手段を有することを特徴とする記憶制御装置。

【請求項2】 データ処理装置が直接アクセスを行う論理的記憶装置と複数にデータを記憶する物理的記憶装置とを対面付け、前記データ処理装置と前記物理的記憶装置との間のデータ転送を制御する記憶制御装置において、前記データ転送の制御の運用中にデータ処理装置の論理的記憶装置へのアクセス情報を指標として採取するアクセス情報採取手段と、前記指標に基づいて前記論理的記憶装置を前記物理的記憶装置に再配置すると共に再配置先の物理的記憶装置にデータを連続的に格納する論理的記憶装置再配置手段とを有することを特徴とする記憶制御装置。

【請求項3】 請求項2に記載の記憶制御装置において、前記アクセス情報は、前記データ処理装置から前記論理的記憶装置へのアクセス頻度情報を含むことを特徴とする記憶制御装置。

【請求項4】 請求項2または請求項3に記載の記憶制御装置において、前記アクセス情報は、前記データ処理装置から前記物理的記憶装置へのアクセス頻度情報を含むことを特徴とする記憶制御装置。

【請求項5】 請求項1に記載の記憶制御装置において、前記指標が、前記論理的記憶装置に求められる信頼性であることを特徴とする記憶制御装置。

【請求項6】 請求項1から請求項5のいずれかに記載の記憶制御装置において、前記指標を保守員に提示する指標提示手段と、保守員からの再配置指示を受け付ける再配置指示受付手段とを具備したことを特徴とする記憶制御装置。

【請求項7】 請求項1から請求項5のいずれかに記載の記憶制御装置において、データ処理装置からの再配置指示を受け付ける再配置指示受付手段を具備したことを特徴とする記憶制御装置。

【請求項8】 請求項1から請求項5のいずれかに記載の記憶制御装置において、前記指標に基づいて再配置の可否を決定する再配置可否決定手段を具備したことを特徴とする記憶制御装置。

【請求項9】 請求項1から請求項8のいずれかに記載の記憶制御装置において、再配置中の論理的記憶装置にデータ処理装置からのアクセスがあったとき、再配置中の論理的記憶装置の再配置完了領域と再配置未完了領域とを識別し、前記アクセス位置が前記再配置完了領域ならば再配置先の論理的記憶装置にアクセスさせ、前記アクセス位置が前記再配置未完了領域ならば当該論理的記憶装置にアクセスさせる。

セス位置が前記再配置完了領域ならば当該論理的記憶装置にアクセスさせるアクセス位置切り換え手段をさらに具備したことを特徴とする記憶制御装置。

【発明の詳細な説明】

05 【0001】

【発明の属する技術分野】 本発明は、記憶制御装置に関する。さらに詳しくは、シケンシャルアクセスの場合やランダムアクセスでヒット率が低い場合でもアクセス性能を向上することが出来る記憶制御装置およびデータの信頼性を向上することが出来る記憶制御装置に関する。特に、本発明は、ディスクレイアウトの高速化ディスク装置、その高速化ディスク装置とディスク制御装置とにより構成される記憶装置サブシステム、およびその記憶装置サブシステムとデータ処理装置とにより構成される情報処理システムに有用である。

15 【0002】

【従来の技術】 シカゴのイリノイ大学で開かれた「ACU SIGMOD」会議において発表された論文「D. Patterson, G. Gibson, and R. H. Kartsis: A Case for Redundant Arrays of Inexpensive Disks (RAID), ACU SIGMOD Conference, Chicago, IL, June 1988, pp. 109-116」は、ディスクレイアウト上のデータ配置に関する技術を開示している。

【0003】 また、特開平7-84732号公報では、ディスク装置の一節をディスクキャッシュの如く用いる技術が開示されている。具体的には、ディスク装置を同時にデータ格納するテンポラリ領域と最終的にデータを書き込む領域とに分け、更新データはバリエーションを生成し、更新データは二重書きし、非同期にバリエーションを生成し、最終領域に書き込む。

【0004】 一方、電気情報通信学会技術研究報告「D E 95-68 (藤木他: Hot Mirroring を用いたディスクアレイのディスク故障時の性能評価、1995年12月、電気情報通信学会技術 Vol. 95-No. 407, pp. 19-24)」には、アクセス頻度の高いデータを保持するRAIDレベルを動的に変更する技術が開示されている。具体的には、ディスク装置をRAID1構成の部分とRAID5構成の部分に分け、ライトアクセスのあったデータを優先的にRAID1構成の部分に格納することにより、データ格納の信頼性を向上させることとなる。

【0005】 また、RAID1構成の部分に格納し、アクセス頻度の低いものはRAID5構成の部分に格納することによる。この技術によれば、記憶容量の異なる物理ディスク装置やRAIDレベルの異なる物理ディスク装置を記憶装置サブシステム内で混在させることが可能であり、論理ディスク装置内のデータを、そのアクセス頻度やアクセスパターンなどの指標に基づいて、任意の物理ディスク装置に格納することが出来る。

【0006】 また、アクセス頻度の高いデータを、より高速な物理ディスク装置に格納するように、動的に格納位置を変更することもある。なお、RAID1のディスクアレイ

【0007】 また、アクセス頻度の高いデータを、より高速な物理ディスク装置に格納するように、動的に格納位置を変更することもある。なお、RAID1のディスクアレイ

とから構成される。

【0025】前述SVP111は、アクセス情報500の保守員への提示や、保守員からの再配置指示620の入力の受け付けを行う。また、保守員からの情報処理システムへの指示の発信や、情報処理システム1の障害状態等の保守員への提示を行う。

【0026】図2は、論理ディスク装置200と物理ディスク装置105の関係を表わした図である。論理ディスク装置200は、データ処理装置100のCPU101が直接アクセスする見掛け上のディスク装置で、アクセスがエラーが実際に格納される物理ディスク装置105と対応している。論理ディスク装置200上のデータは、シーケンシャルアクセスを考慮して、物理ディスク装置105上に連続的に配置されている。論理ディスク装置200のデータが格納されている物理ディスク装置105がディスクアクセス頻度の高い場合、論理ディスク装置200は複数の物理ディスク装置105と対応する。また、物理ディスク装置105の容量が論理ディスク装置200より大きく、複数の論理ディスク装置のデータを上の物理ディスク装置105に格納できる場合には、物理ディスク装置105は複数の論理ディスク装置200と対応する。この論理ディスク装置200と物理ディスク装置105の対応は前述論理物理対応情報300で管理される。例えば、データ処理装置100のCPU101が論理ディスク装置200のデータ201をリードする時、記憶制御装置104で論理物理対応情報300に基づく論理ディスク装置200に対応する物理ディスク装置105を求め、その物理ディスク装置105の領域内のデータ格納位置202を求め、データ転送を行う。

【0027】図3は、論理物理対応情報300を表わした図である。論理物理対応情報300は、論理ディスク装置200と、物理ディスク装置105の構成情報320とから構成される。前述論理物理対応情報310は、各論理ディスク装置200が配置されている物理ディスク装置105上の領域に関する情報であり、論理ディスク装置200から対応する物理ディスク装置105を求める時に用いられる。一方、前述物理物理対応情報320は、各物理ディスク装置105に配置されている論理ディスク装置200に関する情報で、物理ディスク装置105から対応する論理ディスク装置200を求める時に用いられる。

【0028】前述論理物理対応情報310は、物理ディスク装置グループ311、RAID構成312および開始位置313の組を、論理ディスク装置200の数だけ有している。前述論理物理ディスク装置グループ311は、当該論理ディスク装置200が配置されている物理ディスク装置105を示す情報である。前述RAID構成312は、前述論理物理ディスク装置グループ311のRAIDレベルを示す。前述開始位置313は、当該論理

ディスク装置200が物理ディスク装置105上で配置されている先頭位置を示す。

【0029】前述論理物理ディスク構成情報320は、論理ディスク装置グループ321を、物理ディスク装置105の数だけ有している。前述論理物理ディスク装置グループ321は、当該物理ディスク装置105に配置されている論理ディスク装置200を示す。

【0030】図4は、論理ディスク情報400を表わした図である。論理ディスク情報400は、論理ディスク状態401と再配置完了ポインタ402とを、論理ディスク装置200のデータに格納されている。前述論理物理ディスク状態401は、「正常」「閉塞」「フォーマット中」「所配置中」などの論理ディスク装置200の状態を表わす。前述再配置完了ポインタ402は、前述論理物理ディスク状態401が「再配置中」の時のみ有効な情報で、当該論理物理ディスク装置200の再配置処理を完了しているポインタ402よりも前の領域へのアクセスの場合には、再配置後の物理ディスク装置105へアクセスしなければならぬ。一方、再配置は完了ポインタ402以後の領域へのアクセスの場合には、再配置前の物理ディスク装置105へアクセスしなければならない。

【0031】図5は、アクセス頻度情報500を表わしている。アクセス頻度情報500は、アクセス頻度情報501とアクセスパターン情報502とを、論理ディスク装置200の数だけ有している。このアクセス頻度情報500は、記憶制御装置104、データ処理装置100、SVP111のいずれからも参照することが出来る。前述アクセス頻度情報501は、単位時間あたりの当該論理物理ディスク装置200へのアクセス回数とアクセス頻度管理する。このアクセス頻度情報501は、各論理物理ディスク装置200の中でアクセス頻度の高いもの又は低いものを求める指標として用いられる。前述アクセスパターン情報502は、当該論理物理ディスク装置200へのシーケンシャルアクセスとランダムアクセスの割合を管理する。このアクセスパターン情報502は、シーケンシャルアクセスが多く、よりシーケンシャルな性質の高い物理ディスク装置105に再配置するが望ましい論理物理ディスク装置200を求める指標として用いられる。

【0032】次に、記憶制御装置104の動作を説明する。図6は、記憶制御装置104の動作を詳細に表わした図である。まず、リード/ライト処理時の動作について説明する。ディレクタ106は、通常リード/ライト処理を実行する際、CPU101からチャネル103を経由してCPUからの指示600を受け取る。このCPUからの指示600は、リード（またはライト）対象のレコードが記憶されている論理物理ディスク装置200を指定する指定情報1と、リード（またはライト）対象のレ

コードが記憶されている論理物理ディスク装置200内の位置（トラック、セクタ、レコード）を指定する指定情報2とを含んでいる。ディレクタ106は、物理ディスク装置上のアクセス位置算出処理（810）で、前記CPUからの指示600と論理物理対応情報300とを用いて、物理物理ディスク装置105上でアクセス位置を算出する。この物理物理ディスク装置105上のアクセス位置算出処理（810）については図8を参照して後で詳述する。その後、たとえばリード処理では、算出した物理物理ディスク装置105上のデータ格納位置202のデータをキャッシュメモリ107上に読み上げてデータ201とし、その読み上げたデータ201をチャネル103を通じて主記憶102に転送する。

【0033】次に、アクセス情報500の採取処理について説明する。CPU101からのリード/ライト処理のアクセス時に、ディレクタ106は、アクセス対象論理物理ディスク装置200のアクセス情報500を更新する。アクセス頻度情報501の採取は、例えば、アクセスの度に内部カウンタをカウントアップしていき、一定時間または一定回数のアクセス経過後のアクセス時に、前記内部カウンタからアクセス頻度を算出する。アクセスパターン情報502の採取は、例えば、アクセスの度に内部カウンタにシーケンシャルアクセス回数をカウントアップしていき、一定時間または一定回数のアクセス経過後のアクセス時に、前記内部カウンタからアクセスパターンを算出する。

【0034】次に、再配置指示620を説明する。保守員は、SVP111を通じて提示されたアクセス情報500を参照して、各論理物理ディスク装置200の再配置の必要性を検討することになる。再配置を決定した論理物理ディスク装置200があれば、SVP111を通じて記憶制御装置104に対して再配置指示620を出し、この再配置指示620は、再配置対象の論理物理ディスク装置200を2つ指定する指定情報1-2からなる。保守員が行う操作の内容は、後述する第3の実施形態で図10を参照して説明する論理物理ディスク装置再配置要否決定処理（910）と同様である。

【0035】次に、論理物理ディスク装置再配置処理（630）を説明する。ディレクタ106は、前記再配置指示620を受けて、指定された2つの論理物理ディスク装置200の間で論理物理ディスク装置再配置処理（630）を行う。図7は、論理物理ディスク装置再配置処理630の処理フロー図である。ステップ700では、論理物理ディスク情報400のうち指定された2つの論理物理ディスク装置200の論理物理ディスク状態401を「再配置中」に設定する。ステップ701では、論理物理ディスク情報400のうち指定された2つの論理物理ディスク装置200の再配置完了ポインタ402とを論理物理ディスク装置200の先頭位置に初期化する。ステップ702では、論理物理ディスク情報400のうち指定された2つの論理物理ディスク装

置200の再配置完了ポインタ402をチェックし、全領域の再配置が完了していない場合はステップ703へ進む。完了していればステップ707へ進む。

【0036】ステップ703では、再配置完了ポインタ402が示すデータ位置から再配置処理の1回の処理単位分のデータに対して物理物理ディスク装置105からキャッシュメモリ107上へのデータ転送を行う。ここで、1回の処理単位分のデータ量は、再配置対象の2つの論理物理ディスク装置200の元データ1つに対応する各データRAID5の論理物理ディスク装置200とRAID1の論理物理ディスク装置200の間で行うならば、RAID1の論理物理ディスク装置200の元データ1つに対応するデータ量は「1」であるから、1回の処理単位分のデータ量は、RAID5の論理物理ディスク装置200の元データ1つに対応するデータ量と等しい。また、再配置処理の1回の処理単位分のデータ量は、RAID5の論理物理ディスク装置200とRAID1の論理物理ディスク装置200の元データ1つに対応するデータ量に決定される。

【0037】ステップ704では、再配置対象の各論理物理ディスク装置200の再配置完了ポインタ402がパリティを有するRAIDレベルのものである場合、キャッシュメモリ107上の再配置対象の1回の処理単位分のデータ201に対してパリティを生成する。ステップ705では、キャッシュメモリ107上の再配置対象の1回の処理単位分のデータ201および前記ステップ704で作成したパリティを、再配置先の物理物理ディスク装置105へ書き込む。ステップ706では、1回の処理単位分の再配置完了ポインタ402を進める。そして、前記ステップ702に戻る。

【0038】なお、上記ステップ703、704においては、データおよびパリティは、不揮発性メモリ109にも転送して二重化し、キャッシュ障害によるデータ損失を防ぐ。この理由は、上記ステップ705での書き込み時に、例えば、第1の論理物理ディスク装置200と第2の論理物理ディスク装置200のデータのうちの、第1の論理物理ディスク装置200のデータを物理物理ディスク装置105（元は第2の論理物理ディスク装置200に配置されていた物理物理ディスク装置105）へ書き込んだ段階で障害によりキャッシュメモリ107上のデータがアクセス不能になったとすると、書き込みが終了していない第2の論理物理ディスク装置200のデータが消失するからである（元は第2の論理物理ディスク装置200に配置されていた物理物理ディスク装置105には、上記のように第1の論理物理ディスク装置200のデータが書き込まれてしまっている）。

【0039】ステップ707では、論理物理対応情報300を更新する。すなわち、論理物理ディスク構成情報310と論理物理ディスク構成情報321を生成する。ステップ708では、論理物理ディスク情報400の論理物理ディスク状態401を元の状態に戻し、再配置処理（630）を終了する。

【0040】次に、物理物理ディスク装置アクセス位置算出

